*Proc. Natl. Acad. Sci. USA*
Vol. 86, pp. 9597–9601, December 1989
Neurobiology

# Silicon modeling of pitch perception

(periodicity pitch/auditory system/integrated circuits/neural networks/neural autocorrelation)

JOHN LAZZARO AND CARVER MEAD

Department of Computer Science, California Institute of Technology, M.S. 256-80, Pasadena, CA 91125

*Contributed by Carver Mead, September 5, 1989*

ABSTRACT **We have designed and tested an integrated circuit that models human pitch perception. The chip receives as input a time-varying voltage corresponding to sound pressure at the ear and produces as output a map of perceived pitch. The chip is a physiological model; subcircuits on the chip correspond to known and proposed structures in the auditory system. Chip output approximates human performance in response to a variety of classical pitch-perception stimuli. The 125,000-transistor chip computes all outputs in real time by using analog continuous-time processing.**

Many people can sing in key and in unison with a melody. Perceiving the pitch of a sound is an essential part of this task. The diversity of sounds that evoke a distinct pitch indicates the complexity of human pitch perception. We perceive a pure sinusoid as having a pitch that depends directly on its frequency. A weighted sum of sinusoids with harmonic (integer-related) frequencies, $f$, $2f$, $3f$, $4f$, . . . , evokes a pitch identical to a sinusoid of frequency $f$, even if the sinusoid of frequency $f$ in the sum has a weight of zero. We also perceive a distinct pitch in a stereotypical fashion in response to a sum of sinusoids with arithmetically related frequencies and in response to a sum of time-delayed correlated noise signals.

Explanations of the ability to perceive pitch initially used physiological models of auditory processing. An early explanation, suggested by Helmholtz (1), modeled the cochlea as a resonant frequency analyzer. Models developed in the 1950s by Licklider (2, 3) advanced beyond the auditory periphery, specifying several stations of neural computation in explicit detail. However, two limitations impeded further development of physiological models of pitch perception. Auditory neurophysiology was in its infancy and could offer researchers little evidence with which to judge proposed theories. In addition, computer simulation and circuit modeling of neural systems were both technologically limited; published computational verification of the Licklider model, by Lyon,* did not occur until 1984.

As a result, models developed in the 1970s (4, 5) were abstract models; the goal of the research was the description of algorithms, computed by an unspecified "central processor," that exactly matched psychophysical pitch-perception data. These studies contributed essential insights into pitch perception; however, they did not address the implementation strengths and constraints of neural systems.

Advances in auditory physiology and computational neuroscience in the last decade encourage us to return to physiological models of pitch perception. Recent physiological studies have provided insights into the structure and function of both the auditory periphery (6–9) and the auditory brainstem nuclei (ref. 10 and I. Fujita and M. Konishi, personal communication). Also, the tools of computational neuroscience are improving. Integrated-circuit design techniques support the creation of neural models with several hundred thousand computational elements; these models compute neural responses in real time by using analog continuous-time processing (11, 12).

We have designed and tested a silicon integrated-circuit model of pitch perception. The chip receives as input a time-varying voltage corresponding to sound pressure at the ear and produces as output a map of perceived pitch. The chip is a physiological model; subcircuits on the chip correspond to known and proposed structures in the peripheral auditory system and in the auditory brainstem nuclei. The algorithms of the chip share many details with the work of Licklider (3); the chip is an analog integrated-circuit implementation of the work of Lyon*, who proposed computational experiments with the Licklider model and published computer simulations of the performance of the model. The chip output approximates human performance on a variety of classical pitch-perception stimuli.

## System Architecture

Fig. 1 is a block diagram of the chip. The chip receives as input a time-varying signal, corresponding to the sound pressure at the ear. This input connects to a silicon model (13) of the mechanical processing of the cochlea, the organ that converts the sound energy present at the eardrum into the first neural representation in the auditory system. In the cochlea, sound is coupled into a traveling-wave structure, the basilar membrane, which converts time-domain information into spatially encoded information by spreading out signals in space according to their time scale (or frequency). The cochlea circuit is a one-dimensional physical model of this traveling-wave structure; in engineering terms, the model is a cascade of second-order sections with exponentially scaled time constants.

In the cochlea, inner hair cells contact the basilar membrane at regular intervals, converting basilar-membrane movement into a graded half-wave-rectified electrical signal. Spiral-ganglion neurons connect to each inner hair cell, producing action potentials in response to inner-hair-cell electrical activity. The temporal pattern of action potentials encodes the shape of the sound waveform at each basilar-membrane position. Spiral-ganglion neurons also reflect the properties of the cochlea; a spiral-ganglion neuron is most sensitive to tones of a specific frequency, the neuron's characteristic frequency. Axons from spiral-ganglion neurons form the auditory nerve, which carries the first neural representation of audition to the brainstem.

In our chip, inner-hair-cell circuits connect to taps at regular intervals along the basilar-membrane circuit. The inner-hair-cell circuits compute signal-processing operations (half-wave rectification and nonlinear amplitude compression) that occur during inner-hair-cell transduction. Each

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: VLSI, very large-scale integrated.

*Lyon, R. F. (1984) Proceedings, 1984 IEEE International Conference on Acoustics, Speech, and Signal Processing, March 1984, San Diego, CA.

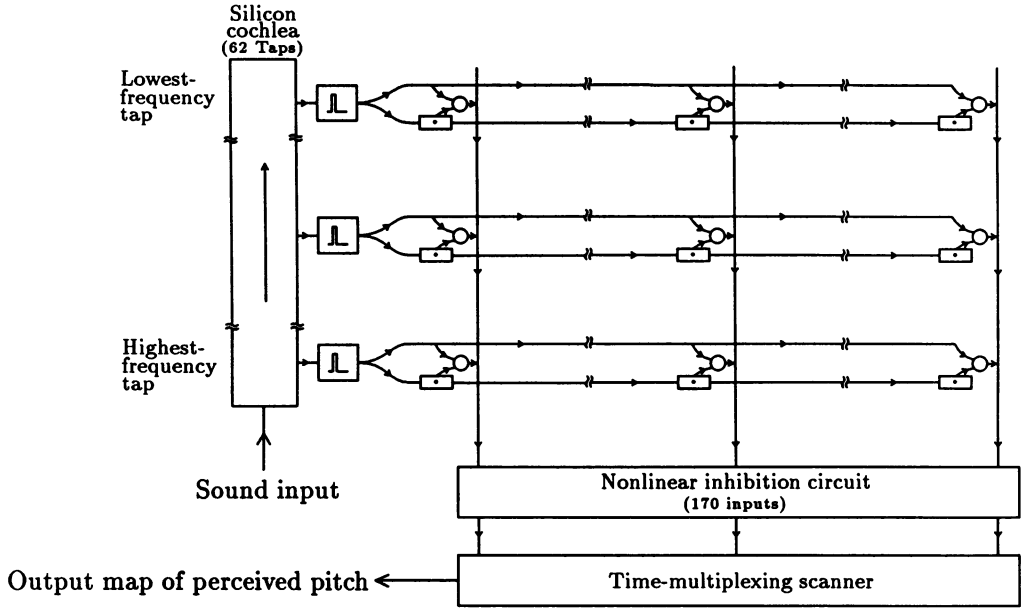


FIG. 1. Block diagram of the pitch-perception chip. Sound enters the silicon cochlea at the lower left of the figure. Circuits that model inner hair cells and spiral-ganglion neurons tap the silicon cochlea at 62 equally spaced locations; square boxes marked with a pulse represent these circuits. Spiral-ganglion-neuron circuits connect to discrete delay lines that span the width of the chip. A small rectangular box, marked with a dot, represents a delay-line section; there are 170 sections in each delay line. A correlation-neuron circuit, represented by a small circle, is associated with each delay-line section. A correlation neuron receives connection from its delay-line section and from the spiral-ganglion-neuron circuit that drives its delay line. Vertical wires, which span the array, sum the response of all correlation neurons that correspond to a specific time delay. These 170 vertical wires form a temporally smoothed map of perceived pitch. The nonlinear inhibition circuit near the bottom of the figure increases the selectivity of this map; the time-multiplexing scanner sends this map off the chip.

inner-hair-cell circuit connects to a spiral-ganglion-neuron circuit. This integrate-to-threshold neuron circuit converts the analog output of the inner-hair-cell model into fixed-width fixed-height pulses. This structure preserves timing information by greatly increasing the probability of pulse events near the zero crossings of the derivative of the neuron's input (14).

The portion of the chip explained thus far models the known structures of the auditory periphery. The remainder of the chip implements proposed neural structures in the brain. In the chip, each spiral-ganglion-neuron circuit connects to a discrete delay line; for each input pulse, a fixed-width fixed-height pulse travels through the delay line, section by section, at a controllable velocity (11). After the circuit has been excited with a single pulse, only one section of the delay line is firing at any point in time. The delay of each section is set not by a global clock, but by a local time constant; due to circuit-element imperfections, section delay times have a spatial standard deviation of about 20% of the mean.

A correlation-neuron circuit is associated with each delay-line section; this circuit receives a connection from the output of its delay-line section and from the spiral-ganglion-neuron circuit that drives the delay line. Simultaneous pulses at both inputs excite the correlation-neuron circuit; if only one input is active, the circuit generates no output. Each row of correlation neurons, associated with a spiral-ganglion neuron, forms a place code of periodicity. A spiral-ganglion neuron fires in a repeating pattern, on average, in response to a periodic signal in the appropriate frequency range. Correlation neurons that fire maximally receive this repeating pattern simultaneously on both inputs; the time delay associated with this correlation neuron is an integer multiple of the period of the signal. In engineering terms, each correlation neuron computes the running autocorrelation function of a filtered version of the sound input for a particular time delay.

In 1951, Licklider (2) proposed this neural autocorrelation structure as a periodicity representation that could be implemented plausibly with synaptic delays in neural circuitry. Although no direct physiological evidence for these autocorrelation structures has been discovered, Carr and Konishi (10) have found direct evidence for cross-correlation structures for auditory localization in the midbrain of the barn owl; these structures use axonal time delays to compute a place code of interaural time delay.

Fig. 2 shows the utility of neural autocorrelation structures in the perception of the pitch of a weighted harmonic sum of sinusoids with frequencies ($f$, $2f$, $3f$, $4f$, . . . ). Due to the filtering action of the cochlea, different sinusoids are predominant in different autocorrelators throughout the chip. Cochlear processing is idealized in Fig. 2; the figure shows an analog representation of the signals in the delay lines across

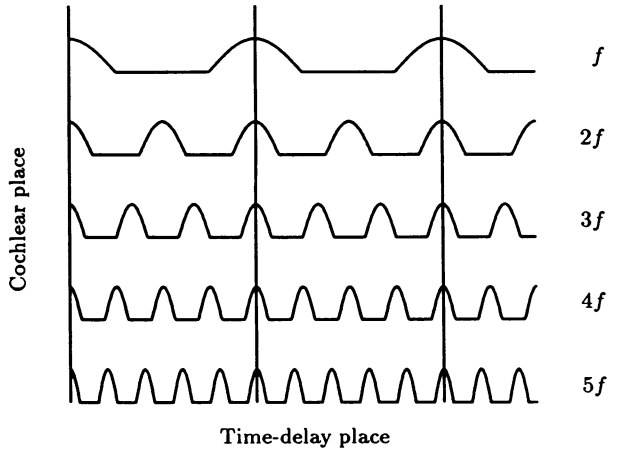


FIG. 2. Analog representation of the signals in the delay lines across the chip in response to a harmonic signal. Cochlear processing is idealized (fully resolved harmonic components, perfect half-wave rectification, no temporal smoothing). Peaks of activity in the horizontal direction coincide with peaks in $f$, shown by vertical lines.

the chip, assuming that all sinusoids are in phase. The peaks in all the delay lines coincide with the peaks of the sinusoid of frequency $f$. Thus, even if the sinusoid of frequency $f$ has zero weight, the representation still encodes the frequency $f$, the perceived pitch of the sum. The outputs of the correlation neurons reflect this representation; in addition, they are invariant to the relative phase of the sinusoids.

To complete his model, Licklider (3) proposed a self-organizing neural network that received connections from the autocorrelation structures and that learned to associate firing patterns with the perception of pitch. For our chip, we designed a simple recognition algorithm suitable for the perception of a single pitch. First, all correlation-neuron outputs corresponding to a particular time delay are summed across frequency channels to produce a single output value. Correlation-neuron outputs are current pulses; a single wire running vertically through the chip acts as a dendritic tree to perform the summation for each time delay.

In this way, a two-dimensional representation of correlation neurons reduces to a single vector; this vector is the map of perceived pitch. The chip then integrates this vector temporally, with an adjustable time constant, providing a stable representation over many cycles of the input signal. Finally, a global shunting-inhibition circuit (15) processes this temporally integrated vector; this nonlinear circuit performs a winner-take-all function, producing a more selective map of perceived pitch. The chip time multiplexes this output map on a single wire for display on an oscilloscope.

## Chip Responses

To show the capabilities and limitations of the silicon model, we recorded chip responses to a variety of classical pitch-perception stimuli. In these experiments, we tuned the basilar-membrane circuit to span about five octaves; lowpass cutoff frequencies ranged from 300 Hz to 10,000 Hz. The delay lines were tuned to provide about 3.3 ms of total delay; with this tuning, the chip perceives pitches above 300 Hz. Temporal smoothing by the recognition algorithm acted with a time constant of tens of milliseconds.

Fig. 3*A* shows maps of a perceived-pitch period generated by the chip in response to sine, triangle, and square waves at various frequencies. As desired, chip response is invariant to the harmonic content of the signal. The chip response shows the first global peak of the autocorrelation representation; the spatial variation in the delay-line timing weakens the strength of subsequent peaks. In Fig. 3*B*, we recorded the map position of the neuron with maximum signal energy for square waves of different frequencies; the graph shows a linear relationship between the input period of the waveform and the pitch period predicted by the chip.

The stimuli in Fig. 4 illustrate the classical "missing fundamental" aspect of pitch perception. Fig. 4*A* shows a narrow-pulse waveform, whereas Fig. 4*B* shows the sum of this narrow-pulse waveform and a synchronized sinusoid with appropriate frequency, amplitude, and phase to cancel exactly the fundamental frequency of the pulse waveform. Human subjects perceive the pitch of both waveforms to be identical (16); Fig. 4*C* shows identical maps from the chip in response to both waveforms at various frequencies.

As in the biological system, the chip circuits that model the cochlear periphery are in some aspects nonlinear and resynthesize the fundamental frequency of the signal in Fig. 4*B*. We have done several experiments to show that the effect of distortion products is negligible. Decreasing the intensity of the stimulus, within the operating range of the chip, does not alter the response map; at lower intensities, spectral analysis of cochlear-circuit outputs shows the strength of the fundamental component of the signal to be near the circuit's noise floor. In addition, chip response does not change when a

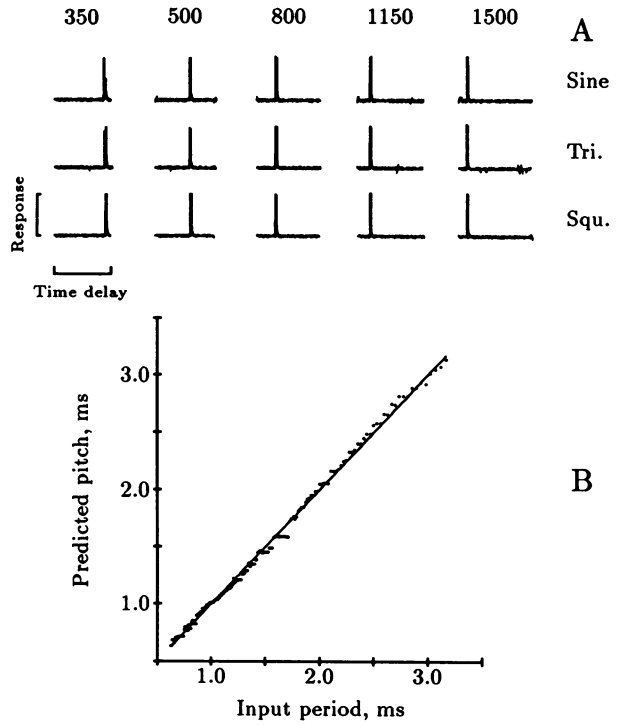


FIG. 3. (*A*) Maps of perceived-pitch period from the chip in response to sine, triangle, and square waves. Column numbers denote frequency in Hz. (*B*) Plot showing map position of the neuron with maximum signal energy for square waves of different frequency; ordinate axis is calibrated from data to indicate pitch period. Dots are data points; solid line shows best linear fit to the data.

lowpass-filtered white-noise signal, with a cutoff frequency above the fundamental of the stimulus, is added to the signal shown in Fig. 4*B* (17).

A sum of three sinusoids, with arithmetically related frequencies $f_c - f_m$, $f_c$, and $f_c + f_m$, is a revealing pitch-perception stimulus; an amplitude-modulated sinusoid, with carrier frequency $f_c$ and modulator frequency $f_m$, as shown in Fig. 5*A*, produces this spectral pattern. If $f_c$ is equal to $nf_m$, where $n$ is an integer, the three sinusoids form an integer-

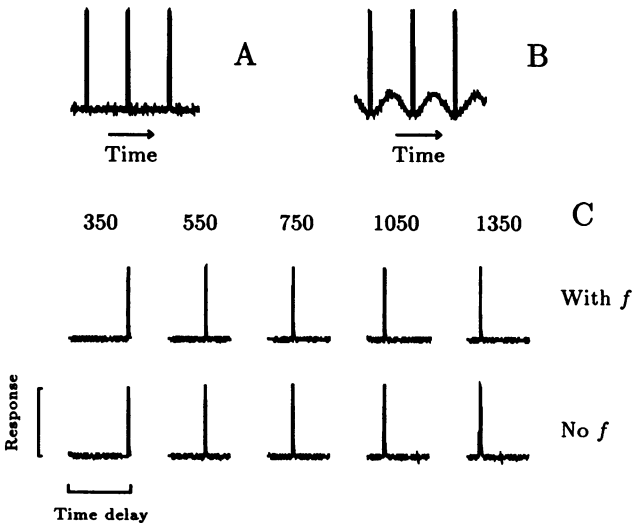


FIG. 4. (*A*) Narrow-pulse sound stimulus. (*B*) Narrow-pulse sound stimulus with canceled fundamental frequency. (*C*) Maps of perceived-pitch period from the chip in response to stimuli shown in *A* and *B* at various frequencies; chip responds identically. Column numbers denote frequency in Hz.

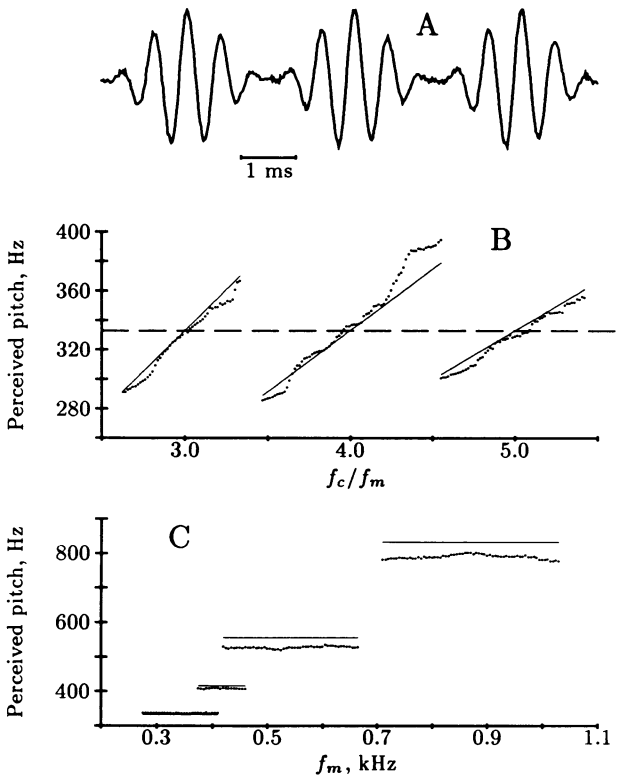


FIG. 5. (*A*) Amplitude-modulated sinusoid sound stimulus. (*B*) Plot showing the center of energy of the chip map in response to stimulus shown in *A*, while the carrier frequency, $f_c$, is varied. Dotted line shows frequency of fixed modulation frequency $f_m$ = 333 Hz. Dots are data points; solid lines show theoretical first-order human response, as explained in text. (*C*) Plot showing the center of energy of the chip map in response to the stimulus shown in *A*, while the modulation frequency, $f_m$, is varied with $f_c$ = 1665 Hz. Dots are data points; solid lines show theoretical first-order human response, as explained in text.

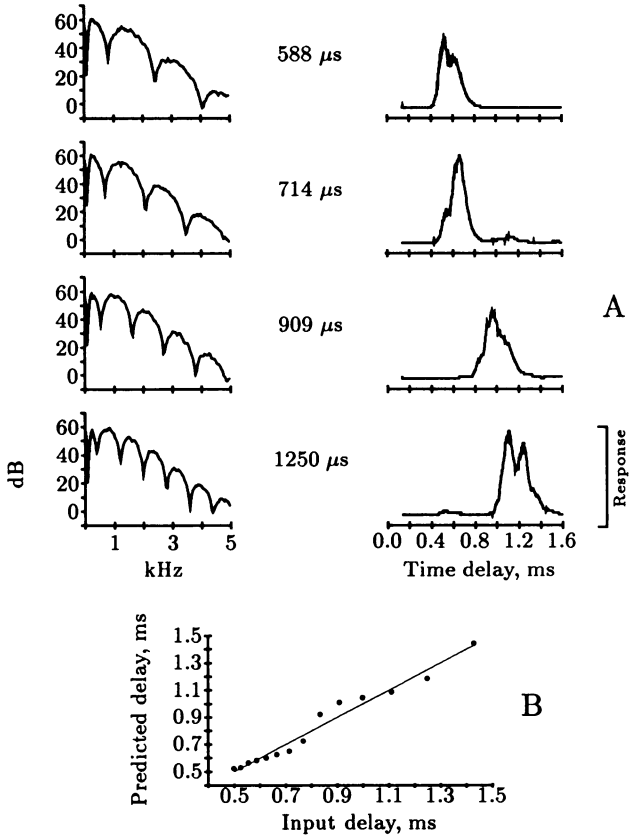


FIG. 6. (*A*) Plots showing maps of perceived-pitch period from the chip in response to a bandpass-filtered sum of two time-delayed correlated noise signals (*Right*). Filtering was kept constant for all plots. Centered numbers indicate time delay between noise signals. (*Left*) Plots show the spectral content of the input stimulus (60-Hz filter bandwidth). (*B*) Plot showing the center of energy of the chip map in response to a bandpass-filtered sum of two time-delayed correlated noise signals, as a function of time delay. Ordinate axis is calibrated from data to indicate perceived delay. Dots are data points; solid line shows best linear fit to the data.

related series, and human subjects perceive a pitch equivalent to that of a sinusoid at the implied fundamental frequency $f_m$. If $f_c$ is equal to $(n + \varepsilon)f_m$, human subjects perceive, to a first order, a pitch equivalent to a sinusoid at the frequency $f_c/n$ (18), where the absolute value of $\varepsilon$ is typically less than 0.5. As postulated by de Boer (18), the human perceptual system calculates a pseudoperiod of this near-harmonic stimulus. The chip response to varying $f_c$, shown in Fig. 5*B*, matches the first-order perception of human subjects.

If $f_c$ is held constant and $f_m$ is varied, human subjects, to a first order, perceive a pitch equivalent to that of a sinusoid with the frequency of the integral submultiple of $f_c$ nearest to $f_m$ (18). The chip response to varying $f_m$, shown in Fig. 5*C*, matches the first-order response of human subjects, limited by the resolution of the output map.

Human perception of amplitude-modulated sinusoids has significant second-order properties. If $f_m$ is held constant and $f_c$ is varied, the perceived pitch is not described exactly by the expression $f_c/n$; the slope of the response is slightly greater than $1/n$. If $f_c$ is held constant and $f_m$ is varied, the perceived pitch is not exactly the integral submultiple of $f_c$ nearest to $f_m$; the perceived pitch decreases slightly with increasing $f_m$ (18). As postulated by de Boer (18), these second-order properties reveal the weighting of individual frequency components in the computation of the pseudoperiod by the human perceptual system. In the chip, the simple recognition algorithm does not support the relative weighting of frequency components; as a result, the responses depicted in Fig. 5 do not show the second-order properties of the human perceptual system.

Human subjects perceive a faint, but distinct, pitch in response to a sum of two time-delayed, correlated noise signals;† the period of the perceived pitch is equal to the time delay. This stimulus is relevant to auditory localization as well as to pitch perception; the outer ear produces time-delayed replicas of incoming sounds that encode the elevation angle of sound sources in mammals (19). Output maps from the chip show a perceived pitch in response to a bandpass-filtered sum of two time-delayed correlated noise signals, as shown in Fig. 6*A*. As shown in Fig. 6*B*, the center of energy of the chip map varies linearly with time delay, in agreement with the linear characteristic of Fig. 3*B*. Like that of human subjects, the response of the chip to the noise stimulus is faint; to obtain the data in Fig. 6, we decreased the integration time constant of the recognition algorithm, and time averaged the responses off-chip.

## Discussion

The chip output approximates human performance in response to a variety of classical pitch-perception stimuli. The major shortcoming of the chip is the inadequate modeling of the second-order properties of pitch perception of amplitude-

†Fourcin, A. J. (1965) in Proceedings of the Fifth International Congress on Acoustics, September, 1965, Liège, Belgium, Vol. 1a, B 52.

modulated tones; this shortcoming is not a failure of neural autocorrelator structures as a representation but rather is a property of the chip's simple recognition algorithm that does not support the relative weighting of frequency components in the recognition process.

Neural autocorrelation structures, at the appropriate time scale, are a natural initial representation for a variety of auditory tasks. Autocorrelation time delays of hundreds of microseconds match the time delays introduced by the outer ear to encode auditory localization information in the elevational plane. Time delays in the millisecond range support pitch perception and complex sound-recognition tasks; time delays of hundreds of milliseconds may form a substrate for the perception of rhythms. As a result, there may be a number of autocorrelation structures in the auditory system at different time scales. Faster delays probably use axonal delay lines, as do the localization structures of the barn owl (10), whereas slower delays probably use neural circuits for delay units. The autocorrelation structures may form a logarithmic map of time, unlike our chip's linear map, to represent compactly many orders of magnitude of time delay.

In conclusion, the pitch-perception chip confirms the practicality of neural autocorrelation structures as a representation of pitch perception in auditory processing. The chip also demonstrates the utility of analog very large-scale integrated (VLSI) circuits as a modeling tool in computational neuroscience. Analog VLSI and neural systems are different in detail, but the frameworks for computation in the two technologies are remarkably similar.

Analog VLSI and neural systems both offer a rich palette of primitives with which to build a structure; nonlinearities are fertile resources for improved system performance. Each subcircuit in the pitch-perception chip is inherently nonlinear, mimicking the nonlinearity of the known or proposed neural structure it models. Analog circuits effortlessly compute nonlinearities in real time.

Analog VLSI and neural systems both pack a large number of imperfect computational elements into a small space. Systems in both technologies must confront these imperfections not as a second-order effect but as a prerequisite for a working design. Random variations between components exist in every subcircuit of the pitch-perception chip. An algorithm implemented with analog VLSI circuits is an algorithm that is robust to component tolerances, an important attribute for a plausible neural model.

Analog VLSI and neural systems are both ultimately limited not by the density of devices, but rather by the density of interconnect. As shown in Fig. 1, the long communication wires in the chip span either the length or the width of the chip but not both; the wire length inside the nonlinear inhibition circuit (schematic not shown) scales linearly with the width of the chip. Analog VLSI technology encourages the creation of models with physiologically realistic connectivity.

These factors illustrate the promise of analog VLSI technology as a modeling tool in computational neuroscience. The issues designers must face in building analog VLSI models suggest properties of neural systems that are difficult to deduce from computer simulation, mathematical analysis, or physiological experimentation.

We thank R. Lyon for valuable contributions throughout the project. We thank R. Lyon, M. Konishi, M. Mahowald, J. Wawrzynek, T. Sejnowski, J. Tanaka, L. Dupré, and D. Gillespie for comments on the research and the manuscript. We thank Hewlett-Packard for computing support and the Defense Advanced Research Projects Agency and the MOS Implementation Service (MOSIS) for chip fabrication. This work was sponsored by the Office of Naval Research and the System Development Foundation.

1. Helmholtz, H. L. F. (1895) *Sensations of Tone*, trans. Ellis, A. J. (Longmans Green, New York), pp. 49–65.
2. Licklider, J. C. R. (1951) *Experientia* **7**, 128–134.
3. Licklider, J. C. R. (1959) in *Pyschology: A Study of a Science*, ed. Koch, S. (McGraw-Hill, New York), Vol. 1, pp. 94–144.
4. Goldstein, J. L. (1973) *J. Acoust. Soc. Am.* **54**, 1496–1516.
5. Wightman, F. L. (1973) *J. Acoust. Soc. Am.* **54**, 407–416.
6. Rhode, W. S. (1971) *J. Acoust. Soc. Am.* **49**, 1218–1231.
7. Kiang, N. Y.-s. (1980) *J. Acoust. Soc. Am.* **68**, 830–835.
8. Kim, D. O. (1984) in *Hearing Science*, ed. Berlin, C. I. (College-Hill, San Diego), pp. 241–262.
9. Dallos, P. (1985) *J. Neurosci.* **5**, 1591–1608.
10. Carr, C. E. & Konishi, M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8311–8315.
11. Mead, C. A. (1989) *Analog VLSI and Neural Systems* (Addison-Wesley, Reading, MA), p. 201.
12. Lazzaro, J. P. & Mead, C. A. (1989) *Neural Computation* **1**, 47–57.
13. Lyon, R. F. & Mead, C. A. (1988) *IEEE Trans. Acoust., Speech, Signal Process.* **36**, 1119–1134.
14. Lazzaro, J. P. & Mead, C. A. (1989) in *Analog VLSI Implementation of Neural Networks*, eds. Mead, C. A. & Ismail, M. (Kluwer, Norwell, MA), pp. 85–101.
15. Lazzaro, J. P., Ryckebusch, S., Mahowald, M. A. & Mead, C. A. (1988) in *Advances in Neural Information Processing Systems 1*, ed. Tourestzky, D. (Morgan Kaufmann, San Mateo, CA), pp. 703–711.
16. Schouten, J. F. (1940) *Proc. K. Ned. Akad. Wet.* **43**, 356–365.
17. Licklider, J. C. R. (1954) *J. Acoust. Soc. Am.* **26**, 945 (A).
18. de Boer, E. (1956) *Nature (London)* **178**, 535–536.
19. Batteau, D. W. (1967) *Proc. R. Soc. London Ser. B* **158**, 158–180.